DOCUMENT RESUME

ED 327 549 TM 015 818

AUTHOR Facione, Peter A.
TITLE The California Critical Thinking Skills Test--College Level. Technical Report #1. Experimental Validation and Content Validity.
PUB DATE 90
NOTE 21p.; For technical reports 2 and 3, see TM 025 819 and TM 015 962.
AVAILABLE FROM California Academic Press, 217 La Cruz, Millbrae, CA 94030 ($10.00; $70.00 for 10).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Tests; College Faculty; *College Students; *Content Validity; Course Evaluation; *Critical Thinking; Higher Education; Instructional Effectiveness; Pretests Posttests; Regression (Statistics); *Skill Development; *Test Validity; Theory Practice Relationship; Thinking Skills
IDENTIFIERS *California Critical Thinking Skills Test (College); California State University Fullerton; Validation Verification and Testing Techniques

ABSTRACT

Four experiments were conducted to determine if the California Critical Thinking Skills Test--College Level (CCTST) measured the growth in critical thinking (CT) skills achieved by college students completing approved CT courses. The experiments, conducted at California State University (Fullerton) involved 1,169 college students in 5 courses with 20 instructors. The theoretical construct for the CCTST is the consensus conceptualization of CT reached by a panel of 46 experts participating in a Delphi research project of the American Philosophical Association during 1987-89. Experiment 1, comparing pretest and posttest scores of students enrolled in CT courses, demonstrated significant growth in CT scores. As a control, the second experiment related CCTST scores of two independent groups enrolled in introduction to psychology courses. The null hypothesis was retained. In the third experiment, using paired pretest/posttest scores, the CCTST measured the growth in CT skills assumed to have occurred as a result of one semester of approved CT instruction. The fourth experiment retained the null hypothesis for the control group using paired pretest/posttest scores. Data are summarized in one table. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

PETER A. FACIONE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED327549

The California Critical Thinking Skills Test -- College Level

Technical Report #1

Experimental Validation and Content Validity

Peter A. Facione
Santa Clara University

c. 1990
CALIFORNIA ACADEMIC PRESS
217 LaCruz Ave., Millbrae , CA 94030

(ERIC Doc. No: TM 015818)

TM 015818

The California Critical Thinking Skills Test: College Level

Technical Report #1 -- Experimental Validation and Content Validity

by

Peter A. Facione

Santa Clara University

Abstract

Technical Report #1 presents the findings of four experiments to determine if the "California Critical Thinking Skills Test: College Level," (CCTST) measured the growth in critical thinking skills achieved by college students completing approved critical thinking courses. Conducted at California State University, Fullerton during the 1989/90 academic year, these four experiments involved 1169 college students, five courses, three departments, 20 instructors, and 45 sections. The theoretical construct grounding the CCTST is the consensus conceptualization of critical thinking articulated by the panel of 46 national experts who participated in a Delphi research project conducted during 1987-1989 for the American Philosophical Association. The CCTST targets five cognitive skills as defined in that Delphi research: interpretation, analysis, evaluation, explanation, and inference. The theoretical construct for the CCTST is directly compatible with the conceptualization of CT promulgated by the California State University System. The CCTST reports six scores: an overall score on CT cognitive skills and five sub-scores named analysis, evaluation, inference, deductive reasoning and inductive reasoning. The first experiment compared the pretest and posttest means for two independent groups of CT students enrolled in 39 sections of four different campus approved CT courses. The CCTST succeeded in detecting the statistically significant growth in CT skills hypothesized to have resulted from CT instruction. As a control, the second experiment related CCTST score of two independent groups enrolled in six sections of introduction to philosophy. The null hypothesis was retained. In the third experiment, using paired pretest/posttest scores, the CCTST measured the growth in CT skills assumed to have occurred as a result of one semester of approved CT instruction. The fourth experiment retained the null hypothesis for the control group using paired pretest/posttest CCTST scores. Generalizing the results, with a confidence interval of 95%, the range of the mean improvement in the CCTST scores of college students completing approved lower division general education CT courses at public comprehensive universities will be bounded by +1.9071 and +.9861. Regression analyses and correlations with GPA, SAT scores, Nelson-Denny Reading Test scores, and other standard measures of academic preparation or ability are presented in Technical Report #2. That report also discusses instructor-related factors, such as CT teaching experience, and the impact of English language ability on the CCTST. Technical Report #3 discusses student-related factors such as academic major, CT self-esteem, gender, and ethnicity. Technical Report #4 provides group norms for the CCTST overall score and for its five sub-scores.

The California Critical Thinking Skills Test: College Level
Technical Report #1 -- Experimental Validation and Content Validity

by

Peter A. Facione

Santa Clara University

This paper reports on research to examine experimentally the validity of the California Critical Thinking Skills Test -- College Level, (CCTST). Published by the California Academic Press, the CCTST is an English language multiple-choice educational assessment tool specifically designed to assess selected, core critical thinking skills, (Facione, 1990 c). The CCTST targets the cognitive skills of interpretation, analysis, evaluation, explanation, and inference. The CCTST is primarily intended for purposes of evaluating the critical thinking skills of college undergraduates in the context of the baccalaureate degree general education requirements.

Long a theoretical concern of psychologists and educators, the growth of the critical thinking movement at both the K-12 and college levels has raised the issue of adequate assessment strategies into a major focus of recent research, (Beyer, 1987; Bloomberg, 1986; Ennis, 1968, 1984 and 1987; Kearney, 1986; Mojeski and Michael, 1983; Norris, 1986, 1989, and 1990; Norris and Ennis, 1989; Resnick, 1990; Siegel, 1988; Sternberg, 1986; and Stewart, 1987). At the college level the critical thinking curriculum has blossomed from the occasional experimental program or ambiguously conceived

introductory logic course into a sharply focused and rapidly expanding area of curricular development. In many North American colleges and universities courses specifically designed to teach critical thinking are being sponsored by a number of different departments. For example, at California State University Fullerton, where this study was conducted, six courses from five different departments are approved as meeting the campus general education requirement in critical thinking. The existence of a growing number of such courses gives rise to the question of how to adequately assess students' critical thinking skills in the context of a given set of instructional or program outcomes.

In addition to a concern about student assessment, a concern expressed by instructors, accreditation bodies, and legislatures, other questions of educational policy also arise. With some campuses, such as the twenty in the massive California State University, now including a critical thinking course in their general education requirements, faculty leaders and cost conscious administrators are raising questions about placement tests and about entry or exit level proficiency standards in critical thinking. Should students, for example, be permitted the option of demonstrating critical thinking ability by examination rather than solely by satisfactory completion of a designated course? Is there any objective evidence which makes it reasonable or unreasonable to expect the same standards of critical thinking proficiency of students regardless of gender, age, number of college units completed, ethnicity, academic major, or native language? Is standardized critical thinking assessment valid, and, if so, is it feasible and educationally desirable?

Since the desirability question is moot unless the validity and feasibility questions are resolved, the current study, and the research out of which it has grown, is chiefly directed at those issues. While a paper and pencil critical thinking assessment tool which focuses on the skills dimension, particularly a tool using the multiple-choice format, would

be only one piece in a total critical thinking assessment package, constructing and validating such a tool would represent a major step toward achieving resolution of some of the most important theoretical and logistical problems which the burgeoning focus on baccalaureate assessment in general and college level critical thinking instruction in particular have generated.

The Theoretical Construct

The CCTST is based on the consensus conceptualization of critical thinking (CT) which emerged from a two-year Delphi research project sponsored by the American Philosophical Association. The panel of experts for the Delphi project included 46 persons active in CT education, research and assessment. Broadly representative of views from a variety of academic disciplines, these persons worked to identify and to characterize core critical thinking skills and dispositions. The Delphi research findings, published in ***Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction,*** (Facione, 1990 a), are briefly reviewed below.

The Delphi panelists began their analysis of CT by identifying the core elements of CT which might reasonably be expected at the freshman and sophomore general education college level. The consensus conceptualization of CT eventually articulated more than a year later by the Delphi group is richly textured. It is this conceptualization of CT which instructors, regardless of their disciplinary orientation, are strongly encouraged to model. In terms of a single sentence, the Delphi panel articulated its understanding of CT as follows:

> We understand CT to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based.

To clarify the above statement, the Delphi panel immediately offered its description of "the ideal critical thinker." By doing so the panel intend to emphasize the view that to inquire regarding the meaning of "critical thinking" requires that one also ask what characterizes successful critical thinkers. Although it is the cognitive skills dimension of CT which is the chief focus of the CCTST, no CT assessment strategy would be fully adequate unless it also addressed CT's dispositional dimension which is captured in the Delphi panel's characterization of the ideal critical thinker.

> The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit.

The Delphi panel identified six cognitive skills as central to the concept of critical thinking. These were interpretation, analysis, evaluation, explanation, inference, and self-regulation. These are defined in the Delphi report as follows:

(1) **Interpretation**, "to comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures or criteria." Interpretation includes the sub-skills of categorization, decoding significance, and clarifying meaning.

(2) **Analysis**, "to identify the intended and actual inferential relationships among statements, questions, concepts, descriptions or other forms of representation intended to express beliefs, judgments, experiences, reasons, information or opinions." Analysis includes the sub-skills of examining ideas, detecting arguments, and analyzing arguments into their component elements.

(3) **Evaluation**, "to assess the credibility of statements or other representations which are accounts or descriptions of a person's perception, experience, situation, judgment, belief or opinion; and to assess the logical strength of the actual or intended inferential relationships among statements, descriptions, questions, or other forms of

representations." Evaluation includes the sub-skills of assessing claims and assessing arguments.

(4) **Inference**, "to identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses, to consider relevant information and to educe the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, concepts, descriptions, questions, or other forms of representation." Inference includes the sub-skills of querying evidence, conjecturing alternatives, and drawing conclusions.

(5) **Explanation**, "to state the results of one's reasoning; to justify that reasoning in terms of the evidential, conceptual, methodological, criteriological and contextual considerations upon which one's results were based; and to present one's reasoning in the form of cogent arguments." Explanation includes the sub-skills of stating results, justifying procedures, and presenting arguments.[1]

A sixth cognitive skill identified by the Delphi panel, and one which the CCTST does not attempt to address, is frequently referred to in the CT literature as meta-cognition. The Delphi panel called it **Self-regulation**, which it defined as "self-consciously to monitor one's cognitive activities,the elements used in those activities, and the results educed, particularly by applying skills in analysis and evaluation to one's own inferential judgments with a view toward questioning, confirming, validating, or correcting either one's reasoning or one's results." Self-regulation includes the sub-skills of self-examination and self-correction.

There is no argument but that assessment strategies other than multiple-choice testing might be as appropriate, if not more appropriate, for evaluating the kinds of cognitive skills and sub-skills listed. Perhaps the best assessment strategy would be the extended non-obtrusive observation by trained raters as subjects interact in a variety of

natural contexts which call for the interactive use of their CT skills. It also seems intuitive that different skills might be better evaluated in different ways. It might be argued, for example, that explanation is best assessed in the context of writing assignments where college students can present their views along with their reasons. However, evidence of the proper application of many of the sub-skills which lead up to that explanation, namely those listed under inference and evaluation, are seldom well-preserved in the final version of a term paper or essay. By its very nature the essay omits claims considered and judged irrelevant, arguments evaluated as not of sufficient significance to the issues at hand to warrant mention, evidence queried by not used in the final from of the essay, alternatives conjectured but ultimately abandoned, and conclusions drawn but ultimately reconsidered and disregarded. It is not the purpose of this research to argue that the multiple-choice strategy is the most appropriate strategy for the assessment of CT skills, only that it is one valid and effective strategy.

In addition to addressing a consensus view of CT experts regarding the meaning of CT in the baccalaureate curriculum, another important consideration in the development of the CCTST was that it should address the CT objectives identified by the California State University system in Executive Order 338. That document specifies that

> instruction in CT is to be designed to achieve an understanding of the relationship of language to logic, which should lead to the ability to (1) analyze, (2) criticize, and (3) advocate ideas, (4) to reason inductively and deductively, and (5) to reach factual or judgmental conclusions based on sound inferences draw from unambiguous statements of knowledge or belief.[2]

Unlike the Delphi report, the California State University Executive Order does not offer sufficient detail to guide assessment research. However, by an ordinary understanding of their terms, the CSU objectives fall well within the range of the cognitive skills identified in the Delphi study, namely analysis, interpretation, evaluation, explanation, and inference.

The CT Skills Assessment Instrument

The CCTST was constructed using a bank of 200 previously piloted multiple-choice items. Thirty-five items were selected on the grounds of their apparent clarity, level of difficulty and discrimination. On the CCTST items 1-5 target **interpretation**, 6-9 **analysis**, 10-13 **evaluation**, 14-24 **inference**, and 25-35 **explanation**.[3] After examining the item analysis for the CCTST based on its first administration to 480 pretest subjects and the initial 465 posttest subjects, item 26 was dropped for lack of discrimination using the point biserial method. For purposes of this research, subsequent statistical analyses were conducted using only the remaining 34 items.

The CCTST is designed to offer several sub-scores of interest. One set of three sub-scores utilizes the Delphi matrix and, borrowing from that terminology, includes sub-scores in "Analysis", "Evaluation" and "Inference."[4] All 34 items are used, with each being assigned to one and only one of the three sub-categories. Operating on the intuitively plausible assumption that interpretation and analysis are closely related, a sub-score on "Analysis" is generated by grouping questions 1-9. Similarly, by relying on the plausible assumption that skills in evaluation and explanation (as tested in the reactive multiple-choice context) are closely related, a sub-score in "Evaluation" is generated by grouping questions 10-13 with 25, and 27-35. Questions 14 through 24 generate the sub-score on "Inference."

In terms of Executive Order 338 of the California State University, the CCTST sub-score on "Analysis" addresses the analysis objective. The "Evaluation" sub-score addresses objectives of criticizing and advocating ideas to the extent that active sub-skills such as advocacy can be accessed at least indirectly using the multiple-choice format. The

"Inference" sub-score speaks to the objective of reaching conclusions based on sound inferences.

A more traditional way of dividing the CT terrain is in terms of deductive as contrasted with inductive inference. For a number of theoretical reasons, not the least of which is the notorious ambiguity of these terms and the inconsistencies found in their use across different disciplines, the deductive/inductive matrix was not used to design the CCTST. However, to address the one remaining California State University skill objective -- to reason inductively and deductively -- the CCTST offers sub-scores on "Deduction" and "Induction." For purposes of these sub-scores items are regrouped as follows: Items 1, 2, 5, 6, 11-19, 22, 23, and 30 produce the sub-score on "Deduction." Items 10, 11, 20, 21, 24, 25, 27-29, and 31-35 yield a sub-score on "Induction." [5]

The First Experiment: An Independent Pre/Post Test

The goal of the first experiment was to determine if the CCTST was sensitive to the differing CT abilities of college students who have or have not completed an approved college level CT course. Naturally, other mitigating factors relating to the students, their instructors, the course itself, the test environment, etc. which might influence student achievement on such a test instrument would have to be identified and controlled. Nonetheless, if the CCTST is satisfactory as a college level assessment instrument it should be able to detect the growth in CT skills that occurs as a result of completing a college level course specifically designed and taught for the purposes of improving CT. This way of proceeding assumes that CT instruction in approved CT courses is effective. Hence, the null hypothesis for purposes of statistical inference is that the instrument would fail to detect statistically significant differences between students who have and have not

completed an approved college level CT course. The alpha level needed to reject the null hypothesis on a one-tailed test was set at $p < .050$. [6]

The primary experiment was conducted by comparing a pretest group (n=480) of students entering required general education CT courses at the start of the spring 1990 semester (February 1990) with a posttest group (n=465) completing the fall semester 1989 sections of the same courses in (November 1989). In the primary experiment discrete cadres of students were used for the pretest and posttest so as to control for the possible contaminating effects of familiarity of the test instrument itself.

The courses selected for study were Psychology 110 "Reasoning and Problem Solving", Philosophy 200 "Argument and Reasoning," Philosophy 210 "Logic" and Reading 290 "Critical Reading as Critical Thinking." Each is a lower division general education course. Each is taught in sections of roughly 25 to 30 students. In all 18 pretest sections and 21 posttest sections were included in the study. The sections were selected to represent the relative proportion of students enrolled in all sections of these four courses. Together the four courses account for 85% of the instruction in general education approved CT courses at California State University, Fullerton, the remainder being conducted chiefly in Speech Communication 235 "Essentials of Argumentation and Debate." With respect to age, gender, college units completed, and ethnicity the samples were determined to be representative of the campus population enrolled in approved CT course.

In all, 945 students comprised the combined Feb. '90 pretest and Nov. '89 posttest groups. Of these, 47.2% were males and 52.8% females (N's = 438 males, 490 females, and 17 cases missing data). In all, 180 students (19.1%) reported that some language other than English was their native language, 761 (80.9%) regard English as their native

language, and in 4 cases these data were missing. Descriptive statistics on eleven other factors help characteristics this student group. (As indicated, cases with data missing were eliminated.)

	Factor	Mean	Std Dev	Minimum	Maximum	Cases
1.	Age in Years	22.4	5.05	17	55	940
2.	BS/BA Sem. Units Complete	71.0	37.06	0	170	877
3.	College GPA (if > 0 units)	2.70	.59	0.0	4.0	877
4.	High School GPA	2.29	1.44	0.0	4.0	877
5.	SAT verbal score	417.4	95.27	200	700	608
6.	SAT math score	484.7	97.14	220	800	608
7.	HS Semesters Prep English	7.72	.97	0	11	583
8.	HS Semesters Prep Math	6.41	1.68	0	12	585
9.	HS Semesters Prep Science	3.88	1.67	0	8	242
10.	HS Semesters Foreign Lang.	1.71	2.43	0	9	529
11.	Self-reported ethnicity:					

CSU ETHNIC CODE SELF-IDENTIFIER

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
American Indian/Native Am.	1	1	.1	.1	.1
Black/Non-Hispanic	2	25	2.6	2.9	3.0
Chicano/Mexican American	3	73	7.7	8.5	11.5
Central American	4	2	.2	.2	11.8
South American	5	7	.7	.8	12.6
Other Hispanic	6	18	1.9	2.1	14.7
Chinese	7	6	.6	.7	15.4
Japanese	8	8	.8	.9	16.3
Korean	9	4	.4	.5	16.8
Southeast Asian	10	8	.8	.9	17.7
Other Asian	11	124	13.1	14.4	32.1
Pacific Islander	12	6	.6	.7	32.8
White/Non-Hispanic	13	533	56.4	62.0	94.9
Filipino	14	15	1.6	1.7	96.6
Other	15	17	1.8	2.0	98.6
Declines to State	17	12	1.3	1.4	100.0
	.	68	7.2	MISSING	
No Response/Missing	16	18	1.9	MISSING	
	TOTAL	945	100.0	100.0	

In terms of their motivation for enrolling in the CT courses, 88% of the students (422 of the 480 pretest group and 411 of the 465 posttest group) indicated that their "main reason for enrolling in the CT course was that it met a campus general education

requirement." It is not unreasonable to assume that the samples are sufficiently large and diverse so as to be a fair representation of the general population of students enrolled in lower division general education in public comprehensive universities throughout the country.

Sections were selected so as to control for a number of factors which might have an affect on how a group of students performs. Such factors as the time of the day or days of the week, for example, might select out students of particular kinds. Four different courses were tested because students from different majors might tend to cluster in different courses (as it turns out they did). Testing conditions, such as the rationale and instructions given students, the time permitted to complete the test, and the quality of discipline in the classroom during the experiment were held constant.

To minimize differences among various sections in the November 1989 posttest students were given no advanced notice of the testing date and were blind as to the exact purpose of the experiment. They were told vaguely that their cooperation was appreciated as part of a much larger university research effort regarding CT. They were told specifically that their individual test results would not affect their final grades. Similar precautions were taken to equalize the motivation with regard to the February 1990 pretest. However, to this investigator, who administered the Feb. pretest and the Nov. posttest to over 80% of the sections in this study, there was an evident difference in the motivational level of the two groups. The Feb. pretest students, perhaps eager to petition into closed courses or generally start the new semester well, seemed more cooperative and appeared to put forth a stronger effort on the CCTST. The November posttest students, pressed at the end of the semester with a variety of deadlines and knowing that the CCTST would not influence their final course grade, although willing to participate, seemed to do hastier work and put forth less effort on the CCTST. If anything, these

differences would tend to minimize if not neutralize whatever gains might otherwise be registered on the posttest over the pretest.

Professors in a given discipline might have different conceptualizations of CT, use different pedagogical approaches, teach from different materials, emphasize different aspects of CT, or be more or less effective in meeting their instructional goals. To best simulate the diverse ways in which CT might be presented and taught by different faculty members in different disciplines and at different universities, 20 faculty persons at various stages in their careers and at different points in their personal reflections about CT and CT pedagogy were involved in this research as instructors. These instructors were selected from among those assigned by their departments to teach these courses. Although they were told that the experiment was intended to validate a CT test, they were not informed, except in very general terms, about the conceptualization of CT which was to be used in the CCTST. They were not permitted to examine copies of the CCTST prior to its administration in their courses as a posttest instrument. And, no attempt was made, other than by virtue of the campus curricular approval process, to standardize, in any way, the syllabi, textbooks, handouts, homework assignments, handouts, our teaching strategies employed by the various instructors. In these respects the experimental situation reasonably approximates the variations in CT instruction and pedagogy one can expect to find throughout the CSU and American higher education today.

The mean number of correct answers out of 34 on the February 1990 pretest was 16.0938 with a standard deviation of 4.654 and a standard error of .212. For the November 1989 posttest the mean 462 was .74 greater at 16.8344, with a standard deviation of 4.678 and a standard error of .217. In both cases the range was 27. The reliability coefficient (KR 20) for the pretest was .69 and for the posttest .68.[7] The resulting t-statistic is 2.44, which, for the one-tailed test, is statistically significant at $p < .0075$. We can be more than

99% confident that the null hypothesis is false -- it is extremely unlikely that the observed difference between the pretest and posttest groups happened by mere chance. This result partially confirms that the CCTST is valid. In other words, given the assumption that the teaching in approved CT courses was effective, the CCTST is sensitive enough to detect the increase in CT skills which resulted. Extrapolating from the samples to the population of general education college students at public comprehensive universities, with a confidence interval of 95% the boundaries of mean improvement evident on the CCTST appears to be +.8473 and +.6339. Given the motivational factors mentioned above, these bounds may, in fact, be too low.

The Introduction to Philosophy Control Group

In a second simultaneous experiment a control group of three sections of Introduction to Philosophy were used. In Nov. '89, 126 students took the CCTST under the same controlled conditions as obtained in the Nov. '89 posttest of the four CT courses. In Feb. '90, 124 students from three sections of Intro. Phil. were pretested using the CCTST. In both the fall and the spring two of the sections were small (25 students) and one was large (80 students). The Feb. '90 pretest mean was 15.436 and the Nov. '89 posttest mean was 15.476 revealing a gain of +.04. The t-statistic for this experiment was .08 and the null hypothesis, that there was no significant difference between the two groups, was retained with P=.938. This suggests that whatever growth in CT skills may have occurred in Introduction to Philosophy, it was not measurable on the CCTST. It also suggests that the gain evidenced in the Nov. '89 CT sections was not the result merely of happenstance or of enrolling in a general education course in a comparable or related discipline.

The Third Experiment -- Paired Samples

In the original Nov. vs. Feb. experiment, separate cadres of students were used for the pretest and posttest samples. This strategy was adopted to control for possible effects of familiarity with the CCTST instrument. However, this strategy created questions of experimental mortality. One concern was that weaker students might have self-selected out of the experiment by having dropped their CT course earlier in the semester. Another concern was that larger numbers of weaker students might have skipped class on the posttest day, since absenteeism in general is much higher in the last weeks of a semester. (To control to some degree the tendency of students to skip class if the time was being spent on an activity which did not affect their final course grade, students were not informed in advance that they would be asked on a given day to sit for the 45 minute CCTST examination.)

In response to the above mentioned concerns a third experiment gathered posttest data in May '90. At that time the CCTST was again administered to those same sections of Psychology 110, Philosophy 200, Philosophy 210 and Reading 290 which participated in the Feb. '90 pretest. Also students in the three Intro. Philosophy control group sections were given the CCTST as a May posttest. To avoid complications arising from instrumentation changes, the identical form of the CCTST was used. To attempt to bring student motivation up to the level apparent during the Feb. pretest, the professors of record were asked to remain in the classroom with the test administrator during the May posttest session. In all other respects the testing situation was essentially the same as had been the case in Nov. '89 and Feb. '90.

In all 323 CT students took both the Feb. '90 pretest and the May '90 posttest. However 61 cases were from two sections of CT taught by this investigator. For a variety

of reasons relating to the possible contamination of the experimental validation study, these 61 cases were withdrawn from subsequent analyses. The remaining 262 cases were examined using a paired t-test analysis. For these 232 cases the pretest mean was 15.9427 with a standard deviation of 4.501 the posttest mean was 17.3893 with a standard deviation of 4.589. The difference + 1.45 was statistically significant at the $P < .000$ level (t-statistic = 6.06). This result indicates that the null hypothesis should be rejected. The CCTST again measured the gain in CT which occurs during one semester of CT instruction. With a confidence interval of 95% we can expect the mean improvement on the CCTST from pretest to posttest to be bounded by +1.9071 and +.9861 in the population of general education college students completing critical thinking instruction at public comprehensive universities.

To further confirm these results, the May '90 posttest mean of 17.3893 was compared to the Nov. '89 posttest mean of 16.8344. For the 262 cases involved the t-statistic was 1.96. This t-statistic was not statistically significant.[8] Hence, in both semesters students who completed an approved CT course did significantly better on the CCTST as compared to those who were only beginning their CT course. No statistically significant difference was found between those who completed their course in the fall and those who completed their CT course in the spring.

The Related Pairs Control Group Experiment

A fourth experiment compared the May '90 posttest score for the Intro. Phil. control group to each student's Feb. '90 pretest scores. The May '90 posttest mean for the control group was 16.36 as compared to a Feb. '90 pretest mean score of 15.72. For the 90 control group students who completed both the Feb. pretest and the May posttest, the

difference is not statistically significant. Comparing the May '90 mean with the Nov. '89 posttest mean of 15.47, we again find no statistically significant difference.

In view of the outcome of the both of the control group experiments, the claim that CT is a naturally occurring by-product of good college instruction seems doubtful. The control group courses were selected because they were generally regarded as solid offerings by more than competent faculty. These colleagues expected improvement in CT skills to be part of what would naturally result from the students' experiences with the kinds of questions discussed and kinds of teaching strategies normally employed in introductory philosophy courses.

Conclusion

We can be confident that the CCTST succeeds in detecting the growth in CT skills which is hypothesized to occur during college level instruction specifically designed for the purpose of critical thinking development. The next questions to ask are (1) How does the CCTST correlate with other measures of academic aptitude and achievement such as GPA and SAT scores? (2) What factors influence the growth of these core CT skills in these specific courses? Regression analyses and correlations with GPA, SAT scores, Nelson-Denny Reading Test scores, and other standard measures of academic preparation or ability are presented in Technical Report #2. That report also discusses instructor-related factors, such as CT teaching experience, and the impact of English language ability on CT skill development as measured by the CCTST. Technical Report #3 discusses student-related factors such as academic major, CT self-esteem, gender, and ethnicity. Technical Report #4 provides group norms and discusses CCTST sub-scores on analysis, evaluation, inference, deductive reasoning and inductive reasoning skills.

Partial Bibliography

Beyer, Barry K., "A Suggested Format for Testing Thinking Skills," *Social Science Record* v24, n1, p3-5, Spr. 1987.

Bloomberg, Fran, *et. al, A Pilot Study of Higher-Order Thinking Skills Assessment Techniques in Science and Mathematics -- Part I Pilot-Tested Tasks -- Part II, Final Report*, National Assessment of Educational Progress, Princeton, NJ, Nov. 1986.

Ennis, Robert H., "Testing for CT: State of the Art," *American Educational Research Association.*, San Francisco, CA, 1968.

_____, "Problems in Testing Informal Logic CT Reasoning Ability," *Informal Logic,*, v6. n1, p3-9, 1984.

_____, "A Bibliography of Testing CT," *CT News*, Center for the Reasoning Arts, CSU Sacramento, v6, n1, Sept.-Oct. 1987.

Ennis, Robert H., and Norris, Stephen P., "CT Testing and Other CT Evaluation: Status, Issues, and Needs," in *Issues in Evaluation*, Algina, James (Eds.), Ablex Press, New York, NY, 1988.

Facione, Peter A., "Assessing Inference Skills," *ERIC Clearinghouse on Tests, Measurement, and Evaluation*, Doc. No: TM 012917, Mar. 1989.

_____, (a) *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*, California Academic Press, Millbrae, CA, 1990; ERIC Clearinghouse on Tests, Measurement, and Evaluation, Doc. No: TM 014423, Feb. 1990.

_____, (b) "Strategies for Multiple Choice CT Assessment," in *CT at Colleges and Universities*, David Hitchcock, (Ed.), Vale Press, Newport News, VA, 1990, forthcoming.

_____, (c) "California Critical Thinking Skills Test: College Level," *California Academic Press*, 217 La Cruz, Millbrae, CA, 94030, Dec. 1990.

Kearney, C. Philip, *et al*, "Assessing Higher Order Thinking Skills," *ERIC Clearinghouse on Tests, Measurement, and Evaluation,*" Apr. 1986.

Kurfiss, Joanne G., *Critical Thinking: Theory, Research, Practice, and Possibilities*, ASHE-ERIC Higher Education Report Number 2, Washington DC, ASHE, 1988.

Modjeski, Richard B., and Michael, William B., "An Evaluation by a Panel of Psychologists of the Reliability and Validity of Two Tests of CT," *Educational and Psychological Measurement*, v43, n4, p1187-97, Winter 1983. [The tests reviewed were the Watson-Glaser CT Appraisal and the Cornell CT Test.]

Norris, Stephen P. "Evaluating CT Ability," *History and Social Science Teacher*, v21, n3, p135-146, Spr. 1986.

_____, "Verbal Reports of Thinking and Multiple-Choice CT Test Design," *Technical Report No. 447*, Champaign, IL: Center for the Study of Reading, University of Illinois, (ERIC Doc. No: ED302826.)

_____, "Effect of Eliciting Verbal Reports of Thinking on CT Test Performance," *Journal of Educational Measurement*, v27, n1, 1990.

Norris, Stephen P., and Ennis, Robert H., *Evaluating CT*, Midwest Publications, Pacific Grove, CA, 1989.

Resnick, L. W., *Education and Learning to Think*, National Academy Press, 1987.

Siegel, Harvey, *Educating Reason: Rationality, CT, and Education*, Routledge, 1988.

Sternberg, Robert J., "CT: Its Nature, Measurement, and Improvement," National Institute of Education, Washington, DC, 1986.

Stewart, B. L., "Testing for CT: A Review of the Resources," *Rational Thinking Reports Number 2*, University of Illinois, Champaign-Urbana, IL, 1987.

Endnotes

(1) Table 4 of the Delphi report provides detailed descriptions and paradigm examples of each sub-skill, (Facione, 1990 a).

(2) The Executive Order goes on to say, "The minimum competence to be expected at the successful conclusion of instruction in CT should be the ability to distinguish fact from judgment, belief from knowledge, and skills in elementary inductive and deductive processes, including an understanding of the formal and informal fallacies of language and thought."

(3)Suggested strategies for framing questions which target the various skills are described in "Assessing Inference Skills" (Facione, 1989), and "Strategies for Multiple-Choice CT Assessment," (Facione, 1990 b).

(4) The terms "analysis" and "evaluation" as used here are broader than as used in the Delphi research. Specifically, the term "analysis" refers to both analysis and interpretation as described in the Delphi study. Likewise the term "evaluation" refers to both evaluation and explanation.

(5) The distinction between induction and deduction is drawn on the basis of the purported strength of the inference. If the inference is such that its conclusion is purportedly necessitated by its premises, the inference is deductive. If the conclusion is porportedly warranted, but not necessitated, the inference is inductive. Because of the conceptual ambiguities associated with the deduction/induction distinction as it operates in different disciplines, there is a great disutility associated with the use of these terms.

(6) Persons unfamiliar with statistical inference notation might find it more intuitive to interpret the alpha level as meaning that the odds that the data should turn out as they did merely by chance are less than 5 in 100. In other words, if this alpha level is met, one could say, with 95% confidence, that one is not declaring false an hypothesis which is, in fact, true.

(7) Norris and Ennis recommend reliability ratings within the .65 to .75 range. Unlike tests which focus on a single skill, "there is no theoretical reason for believing that all the items on [CT tests] should correlate highly with one another... Very high reliabilities, especially on tests purporting to test a variety of aspects of CT should not be considered automatically better than more moderate ones," (Norris and Ennis, 1989, p. 46f).

(8) With 261 degrees of freedom the probability using the two-tailed test was .051.